electronics

Article

Instance Segmentation of Underwater Images by Using Deep Learning

Jianfeng Chen [1,*], Shidong Zhu [2] and Weilin Luo [2]

[1] Pen-Tung Sah Institute of Micro-Nano Science and Technology, Xiamen University, Xiamen 361102, China
[2] College of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China; zsd09232022@163.com (S.Z.); wlluo@fzu.edu.cn (W.L.)
* Correspondence: 19920220156449@stu.xmu.edu.cn

Abstract: Based on deep learning, an underwater image instance segmentation method is proposed. Firstly, in view of the scarcity of underwater related data sets, the size of the data set is expanded by measures including image rotation and flipping, and image generation by a generative adversarial network (GAN). Next, the underwater image data set is finally constructed by manual labeling. Then, in order to solve the problems of color shift, blur and the poor contrast of optical images caused by the complex underwater environment and the attenuation and scattering of light, an underwater image enhancement algorithm is used to first preprocess the data set, and several algorithms are discussed, including multi-scale Retinex (MSRCR) with color recovery, integrated color model (ICM), relative global histogram stretching (RGHS) and unsupervised color correction (UCM), as well as the color shift removal proposed in this work. Specifically, the results indicate that the proposed method can largely increase the segmentation mAP (mean average precision) by 85.7% compared with without the pretreatment method. In addition, based on the characteristics of the constructed underwater dataset, the feature pyramid network (FPN) is improved to some extent, and the preprocessing method is further combined with the improved network for experiments and compared with other neural networks to verify the effectiveness of the proposed method, thus achieving the effect and purpose of improving underwater image instance segmentation and target recognition. The experimental analysis results show that the proposed model can achieve a mAP of 0.245, which is about 1.1 times higher than other target recognition models.

Keywords: underwater image; deep learning; instance segmentation; image enhancement; data augmentation

check for updates

Citation: Chen, J.; Zhu, S.; Luo, W. Instance Segmentation of Underwater Images by Using Deep Learning. *Electronics* **2024**, *13*, 274. https://doi.org/10.3390/electronics13020274

Academic Editor: Chiman Kwan

Received: 26 December 2023
Accepted: 4 January 2024
Published: 8 January 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the process of exploring the ocean, many underwater tasks are completed by underwater robots that can replace human beings to perform some dangerous work [1]. The application of machine vision and underwater image processing technology can help underwater robots work more efficiently in the exploration of marine life [2], underwater archaeology [3], underwater environment detection [4], marine resources exploration [5], marine engineering construction [6], aquaculture [7], etc.

Image segmentation is a classic issue in the field of machine vision research. The so-called image segmentation refers to dividing an image into several specific and non-intersecting regions, which have unique characteristics (e.g., gray scale, geometric characteristics and color, etc.). The characteristics in the same region are consistent or similar. On the contrary, the characteristics of different regions are obviously different. In short, the goal of image segmentation is to separate the target from the background in an image.

Traditional image segmentation algorithms are mostly based on digital image processing, topology, mathematics and other disciplines. Representative algorithms include the algorithms based on threshold, morphology, clustering and region. The threshold-based

image segmentation algorithm is a common method, which uses the differences of different features (e.g., color, gray level, outline, etc.) between different targets to select a threshold to divide the target and the background. Yang [8] establishes a two-dimensional membership matrix based on the domain information of image pixels, and the optimal threshold is obtained by setting the criterion function as a two-dimensional multi-threshold fuzzy divergence. Liu [9] proposed a multi-threshold image segmentation method, which combined the search strategy with a differential evolution algorithm and gray wolf algorithm. Morphology based image segmentation is simple and fast. Ma [10] introduced a morphological filter group and a segmentation method based on texture feature uniformity and a contrast criterion to solve the over-segmentation problem of the watershed algorithm. Clustering based segmentation is to firstly represent the image pixels with feature space points; then cluster them according to their positions; and finally map the results back to the image to achieve segmentation. Mariena [11] proposes a hybrid K-means clustering algorithm with cluster center evaluation, which overcomes the problem that the initial cluster center number and center position have a great influence on the segmentation effect. Cai [12] proposed a fuzzy C-means clustering algorithm, by integrating spatial information, introducing a two-dimensional histogram of the image into the traditional fuzzy C-means clustering algorithm and improving the membership function. The results show that a better segmentation effect and stronger anti-noise performance are achieved.

In the research of underwater image segmentation, Wang [13] proposed an underwater image segmentation algorithm based on an entropy constraint and fast fuzzy C-means clustering by combining the gradient operator, histogram characteristics of image, sampling calculation and relative information loss of an image. By introducing the salient features of underwater images and an edge feature extraction method based on depth information, Sun [14] proposed a level set underwater image segmentation algorithm that integrated region and edge features and standardized the level set function by using the distance regularization term, which enhanced the stability of the evolution of the level set function and achieved good segmentation results for underwater images. Li [15] redefined the fuzzy entropy based on the fuzzy entropy segmentation algorithm, used the improved particle swarm optimization algorithm to search the segmentation threshold and proposed an underwater image segmentation algorithm. The results show that it has a good segmentation effect for underwater images with a simple background and has stronger adaptability and noise resistance. Based on the characteristics of image gray fluctuation, Yan [16] adopted an asymmetric adaptive median filtering method to improve the edge characteristics between the foreground object and the background of the image and achieved good results in underwater image segmentation.

Due to the complexity of the underwater environment, underwater images often degenerate. The effect of segmentation by means of traditional image segmentation technology on underwater targets is not satisfactory. With the rise of machine learning, deep learning has been applied to underwater image segmentation. Image segmentation based on deep learning mainly includes semantic segmentation and instance segmentation. Semantic segmentation is pixel-level image segmentation. Each pixel in the image is divided into corresponding categories to achieve segmentation. Unlike in semantic segmentation, in instance segmentation not only should pixel-level classification be carried out, but also different examples should be distinguished on the basis of specific categories. With the rapid development of deep learning technology, image segmentation based on deep learning has become a research hotspot in recent years. Long [17] proposed a full convolution network, which has no limitation on the size of the input image. It mainly used convolution layers to replace the fully connected layers in the traditional convolutional neural network and applied transposed convolution to the feature map of the last convolution layer for upsampling to ensure that the size of the upsampled feature map is the same as the size of the input image. Finally, the upsampled feature map was classified pixel by pixel, and the semantic level segmentation of the image was achieved. Zhao [18] achieved good results in semantic segmentation by using a pyramid pool network to fuse the contextual information

of images. Yu [19] proposed a hole convolution method, which aggregated the multi-scale context information without reducing the resolution of the image and achieved good results in the dense prediction of semantic segmentation. Chen [20] proposed a DeepLab model, which combined the deep convolution neural network and the probability graph model, used hole convolution to expand the receptive field and obtain more contextual information and adopted a completely connected conditional random field to improve the model's ability to capture details. Dai [21] puts forward the case sensitive score graph based on the full convolution network and used two full convolution integral branch networks. One branch generated all the case sensitive score graphs through the aggregation module, the other branch calculated the object score graph and the two branches combined to realize instance segmentation. He [22] proposed a two-stage detection model, Mask R-CNN. Based on Faster R-CNN [23], this model added a branch to predict the information of the target segmentation mask and performed tasks including target detection, instance segmentation and human key point detection. In the research of underwater image segmentation, Cao [24] used the FCN-DenseNet network as the segmentation framework and the YOLOv5 network as the target detection framework and achieved the multi-objective semantic segmentation of underwater images by integrating two algorithms. Chen [25] put forward a FUSS model with information interaction, in which the extracted global feature information of the query image interacts with the local feature information of the supporting image target and achieves good results in underwater semantic segmentation with few tags. Yue [26] segmented the network based on PSPNet, used ResNet101 as the feature to extract the network, introduced depth separable convolution to reduce the computational complexity and improve the loss function and achieved good segmentation results in underwater fish images. Based on the Mask R-CNN network, Hu [27] has achieved good results in the segmentation of underwater sea cucumber and starfish images by replacing the original backbone network, ResNet, with Swin Transformer, enhancing underwater images by using the Water-Net network and replacing the traditional NMS algorithm with the Soft-NMS algorithm.

Although deep learning based instance segmentation of underwater images has been applied, further studies are needed. First of all, due to environmental limits, underwater image data sets are difficult to obtain. As a result, the segmentation effects based on deep learning might degrade. Secondly, due to light attenuation and scattering, underwater images are of low quality, which affects the segmentation results. Thirdly, the structure of the deep learning network can be modified to improve the accuracy and speed of segmentation. In the study, some measures are taken with respect to these three aspects. Firstly, the size of the underwater image data set is expanded by using image rotation, flipping, and a GAN network. Secondly, an underwater image enhancement algorithm is used to preprocess the data set. Thirdly, to reconstruct the deep learning network, the last layer of the feature map with the largest receptive field in the feature pyramid network (FPN) is removed. In addition, a lightweight feature extraction network, MobilenetV2, is used to replace the original backbone network. The experimental results show that the proposed combination of improved deep neural network and image enhancement under complex underwater environments can significantly improve the segmentation effects of underwater images.

The experimental environment was as follows: Python version 3.6.13, Opencv version 4.6.0.66, and pytorch version 1.10.0. In addition, the computer operating system was Windows 10, and the hardware configuration was as follows: the processor was an Intel core i5-10600kf (6 cores and 12 threads); the graphics card was an NVIDIA GeForce RTX 3080 (10 g); the CUDA version was 11.0; the version of the cudnn (gpu Acceleration Library) was 8.0.4; and the memory module had a capacity of 32 G.

The rest of this paper is organized as follows. In the Section 2, the construction of underwater image data set is introduced, including the selection of the initial image, data expansion (including image rotation and flipping, and the expansion of data sample capacity by images generated by the GAN network) and the labeling of images, and finally a kind of underwater image data set is constructed. In the Section 3, the image enhancement algorithms (including MSRCR, ICM, RGHS, UCM and two combined optimization algorithms based on deep learning) are introduced to preprocess the underwater image and the results of the preprocessing of the underwater image are evaluated. In the Section 4, the adjustment of the Mask R-CNN neural network and the validation of the adjusted network are introduced, as well as the improvement of the network after underwater image preprocessing.

2. Construction of Underwater Image Data Set

In image segmentation, the construction of the data set is the primary task. At present, because the acquisition of underwater images requires professional personnel and equipment, there are few related free underwater image data sets. In addition, due to the complex underwater environment, which contains many impurities and particles, underwater imaging has a series of problems, such as color differences, poor contrast and detail loss, so it is difficult to obtain high-quality underwater data sets. In this study, based on the traditional data augmentation method and GAN network selection, an underwater image dataset for instance segmentation was constructed. The initial images of the data set constructed come from the 2020 underwater robot target capture competition in the Underwater Robot Picking Contest (URPC2020).

The construction process of the underwater image data set is shown in Figure 1. First, image transformation, including rotation and flipping, is utilized to enhance the diversity and variability of the dataset. These transformations help introduce variations in the orientation and spatial arrangement of the images, making the model more robust and capable of generalizing to different scenarios. Second, the ConSinGan network is applied to achieve underwater image data set augmentation. The dotted box in Figure 1b represents the processing methods applied at various stages of the ConSinGan. ConSinGan is a generation countermeasure network, which can train the whole network with one image [28,29]. Limited by data samples and network training time, a model trained with only one image has great application value. As shown in Figure 1b, in the first stage (stage 0), a model with three convolutional layers is applied to generate a low-resolution rough image from random noise. In the second stage (stage 1), after n rounds of training, another three convolutional layers are added to the network to generate a higher resolution image; the input to the second step is computed using the output of the previous step. Then, one repeats the second step until the generation reaches the desired image resolution. At the same time, one should add random noise while training the network for the last three stages. Third, image annotation is performed. The method adopted is a Mask R-CNN network in deep learning. Since it belongs to supervised learning, one needs to use a set of labeled data to learn the mapping relationship from input to output. In order to achieve the segmentation of underwater image instances, the position and category of the target in the image and the mask to which the target pixel belongs are needed. The image annotation tool used is labelImg [30] and labelme [31]. The generated data sets are in visual object classes (VOC) format.

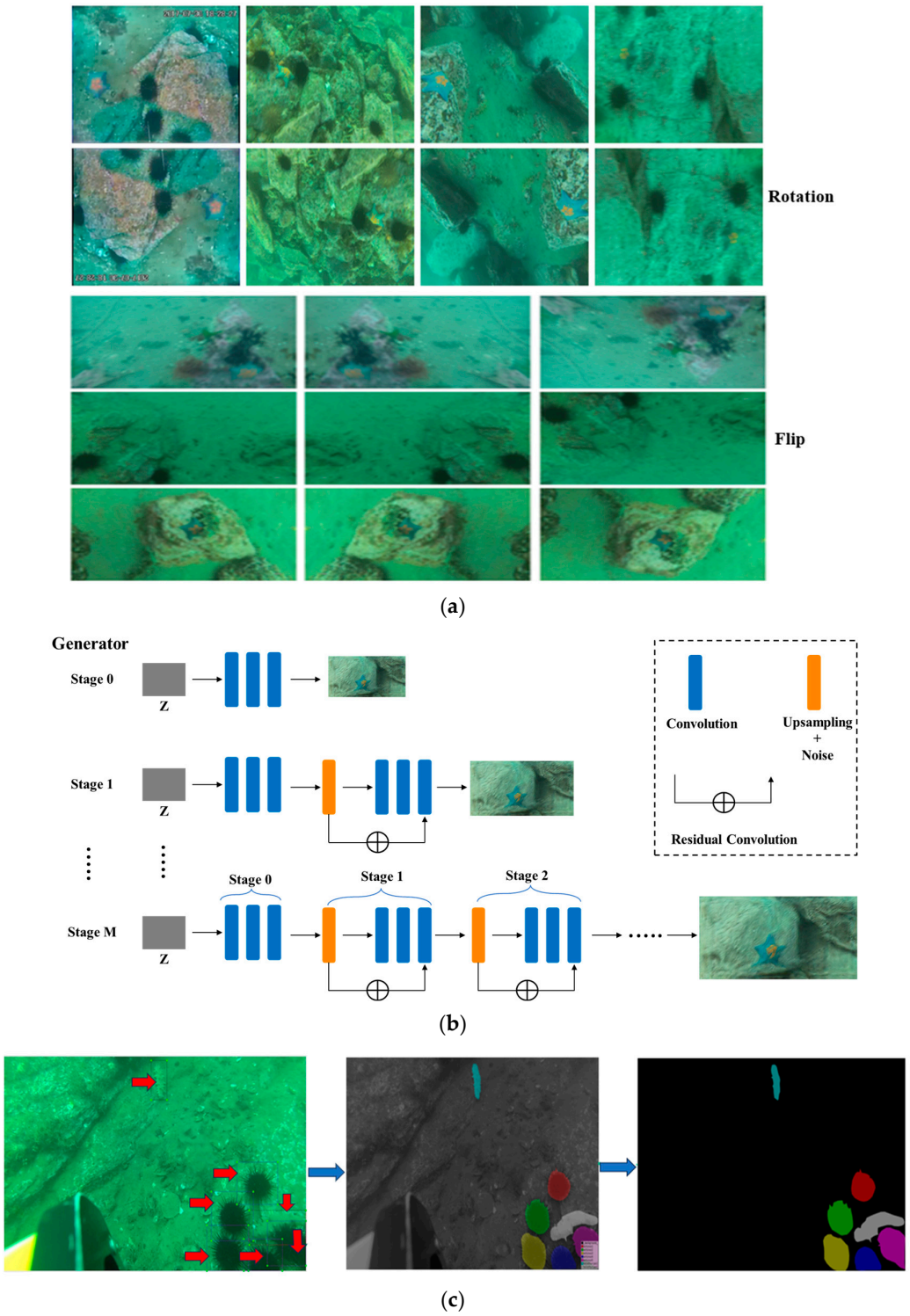


Figure 1. Construction of underwater image dataset. (**a**) Image rotation and flip; (**b**) image generated by ConSinGan; (**c**) image annotation. The red arrows indicate target locations.

3. Image Preprocessing

When light propagates underwater, it will be attenuated and absorbed continuously. At the same time, underwater impurity due suspended particles will scatter the light. As a result, the underwater images present excellent bias, blur, poor contrast and unclear edge information. In order to obtain accurate information in underwater images, it is necessary to preprocess these images. Several classic underwater image enhancement

algorithms are available such as Retinex algorithms [32–35], the integrated color model algorithm (ICM) [36], the relative global histogram stretching algorithm (RGHS) [37], and unsupervised color correction methods (UCM) [38]. In order to comprehensively deal with the degradation of underwater images. We propose a combinatorial optimization algorithm to enhance the underwater image from the aspects of color restoration, dehazing and improvement of contrast.

3.1. Color Cast Removal of Underwater Images

In this work, we adopted two algorithms to remove color deviation. One is an improved white balance, and the other is color channel separation correction of the Lab color model.

(A) Improving white balance

The white balance algorithm is a classic color correction algorithm, which can make the color of the image more natural and realistic. The principle of the white balance algorithm is to adjust the brightness of the corresponding colors so that the overall brightness of the adjusted image is more uniform, and the color distortion of the image caused by uneven illumination can be reduced. Underwater images usually show a blue-green color, which leads to image color distortion. By applying white balance, the color cast can be corrected, so that the color of images in different light sources can be restored to its original color.

Commonly used white balance algorithms include the gray world algorithm and perfect reflection algorithm. For the gray world algorithm, it is regarded that for an image with a significant color change, the average value of the pixel values of R, G and B channels tends to the same value. Based on this assumption, we can calculate the average pixel values of the R, G and B channels, and then obtain the average value of the three channels, which is regarded as a "gray" constant value; Then, the gain coefficient of the corresponding channel can be obtained by comparing the average value with the average value of each channel of the three channels. Finally, the adjusted pixel value of each channel is obtained by calculating the gain coefficient of each channel.

$$\begin{cases} R_{\text{aver}} = \frac{1}{MN}\sum_{\text{i}=1}^{M}\sum_{j=1}^{N} R_{ij} \\ G_{\text{aver}} = \frac{1}{MN}\sum_{\text{i}=1}^{M}\sum_{j=1}^{N} G_{ij} \\ B_{\text{aver}} = \frac{1}{MN}\sum_{\text{i}=1}^{M}\sum_{j=1}^{N} B_{ij} \end{cases}, \tag{1}$$

$$Gary = (R_{aver} + G_{aver} + B_{aver})/3, \tag{2}$$

$$\begin{cases} KR = (R_{aver} + G_{aver} + B_{aver})/(3 * G_{aver}) = Gary/R_{aver} \\ KG = (R_{aver} + G_{aver} + B_{aver})/(3 * G_{aver}) = Gary/G_{aver} \\ KB = (R_{aver} + G_{aver} + B_{aver})/(3 * B_{aver}) = Gary/B_{aver} \end{cases}, \tag{3}$$

$$\begin{cases} R_{\text{new}} = C(\text{R}) \times KR \\ G_{\text{new}} = C(\text{G}) \times KG \\ B_{\text{new}} = C(\text{B}) \times KB \end{cases}, \tag{4}$$

where M and N represent the length and width of the image, respectively. R_{aver}, G_{aver} and B_{aver} represent the average pixel values of R, G and B channels, respectively. KR, KG and KB represent the gain coefficients of R, G and B channels, respectively. R_{new}, G_{new} and B_{new} represent the processed pixel values of the three channels, respectively. C(R), C(G) and C(B) are the channel values in R, G and B channels.

The conventional white balance algorithm can correct the color deviation of the image in an atmospheric environment well. However, it is not ideal for the color deviation correction of underwater images. Therefore, we adopted an improved white balance algorithm [39]. The main principle of this algorithm is to use the channel component with

the least attenuation of light energy, that is, the lightest color distortion, as the reference value to compensate the other two channels. From the previous analysis, it can be seen that in the underwater environment, blue light and green light are the least attenuated, and red light is basically attenuated when it is about 8 m underwater. Therefore, it can be assumed that the attenuation is the blue channel is the least. Firstly, the average gray values of the R, G and B channels are calculated, and then the compensation value is obtained by subtracting the average gray values of the blue channel from the average gray values of the red and green channels. Finally, the pixel values of the red and green channels are adjusted according to the compensation value, and the gray levels of the red and green channels are moved to the gray level positions similar to those of the blue channel, while the blue channel remains unchanged.

$$\begin{cases} d_{B_R} = m_B - m_R \\ d_{B_G} = m_B - m_G \end{cases}, \tag{5}$$

$$\begin{cases} R' = C(\mathrm{R}) + d_{B_R} \\ G' = C(\mathrm{G}) + d_{B_G} \end{cases}, \tag{6}$$

where m_R, m_G and m_B respectively represent the average gray values of R, G and B channels; d_{B_R} and d_{B_G} respectively represent the compensation values of red and green channels; R' and G' respectively represent the pixel values of red and green channels after processing.

(B) Color channel separation and correction of Lab color model

The RGB model is a common image color model. Another optional model is the Lab color model [40]. In the study, color channel separation and the correction of the Lab color model are proposed. The Lab model consists of one brightness channel and two color channels, where L represents brightness; a represents from dark green (low brightness value) to gray (medium brightness value) to bright pink (high brightness value); b represents from bright blue (low brightness value) to gray (medium brightness value) to yellow (high brightness value). To perform the color channel separation correction algorithm, the gray value in the middle position of color channels (that is, channels a and b) is used as the reference value for color correction, and then the compensation value is obtained by subtracting the central value of color channels from the reference value. Thus, the gray value of channels *a* and *b* is compensated by the compensation value to realize color correction, in which the correction of channels *a* and *b* is carried out separately.

Using this algorithm to determine the color shift of underwater images, the first step is to convert RGB images into Lab images. Because there is no direct conversion between the two color model images, it is necessary to make a transitional conversion through the XYZ model, that is, RGB→XYZ→Lab. The conversion procedure is as follows.

A commonly used linear transformation matrix is [41]:

$$M = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix}, \tag{7}$$

Then, the transformation from RGB to XYZ can be expressed as

$$\begin{bmatrix} X^* & Y^* & Z^* \end{bmatrix} = \begin{bmatrix} C(\mathrm{R}) & C(\mathrm{G}) & C(\mathrm{B}) \end{bmatrix} M^T, \tag{8}$$

where X^*, Y^* and Z^* represent the channel values in the XYZ model, respectively.

Transformation from XYZ to Lab can be expressed as

$$\begin{cases} L^* = 116 f(Y^*/Y_n) - 16 \\ a^* = 500[f(X^*/X_n) - f(Y/Y_n)] \\ b^* = 200[f(Y^*/Y_n) - f(Z/Z_n)] \end{cases}, \tag{9}$$

where f () represents nonlinear transformation

$$f(t) = \begin{cases} t^{1/3}, \ t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^3 t + \frac{4}{29}, else \end{cases}, \tag{10}$$

where, L^*, a^* and b^*, respectively represent the values of one brightness channel and two color channels in the Lab model; X_n, Y_n, and Z_n represent the gray values of the reference white point in the XYZ color model in the three channels.

After being converted into a Lab image, the median of the gray values of channel a and channel b are obtained. Then, it is subtracted from the gray values (reference values) in the middle position of each channel to obtain a correction value. Afterwards, it is added to the initial gray values of each channel, respectively to obtain the processed gray values of channel a and channel b, which are expressed as

$$\begin{cases} a' = a^* + 128 - M_a \\ b' = b^* + 128 - M_b \end{cases}, \tag{11}$$

where a'. and b' respectively represent the corrected gray values; M_a and M_b are the median of the gray values; 128 is the reference value. In order to transform the gray scale range of a Lab image into the same as that of an RGB image, that is, from [–128, 127] to [0, 255], the reference value is set as 128. In addition, if the channel gray value is greater than 255, it is set to 255. Otherwise, if it is less than 0, it is set to 0. It is noted that for the outliers, two measures are available, including minimization/maximization and normalization. Usually, minimization/maximization makes an image brighter while normalization makes an image darker. Considering that underwater images are usually dull in color, in this study, the minimization/maximization method was employed.

Figure 2 shows the results by using the improved white balance algorithm (marked as algorithm A) and the color channel separation correction of the Lab color model (marked as algorithm B) to remove the color bias of some underwater data sets. Compared with the original image, it can be seen that both algorithms have good color cast removal effect. Comparatively speaking, method A is better than method B, because the colors of different objects in the image processed by method A are more distinct.

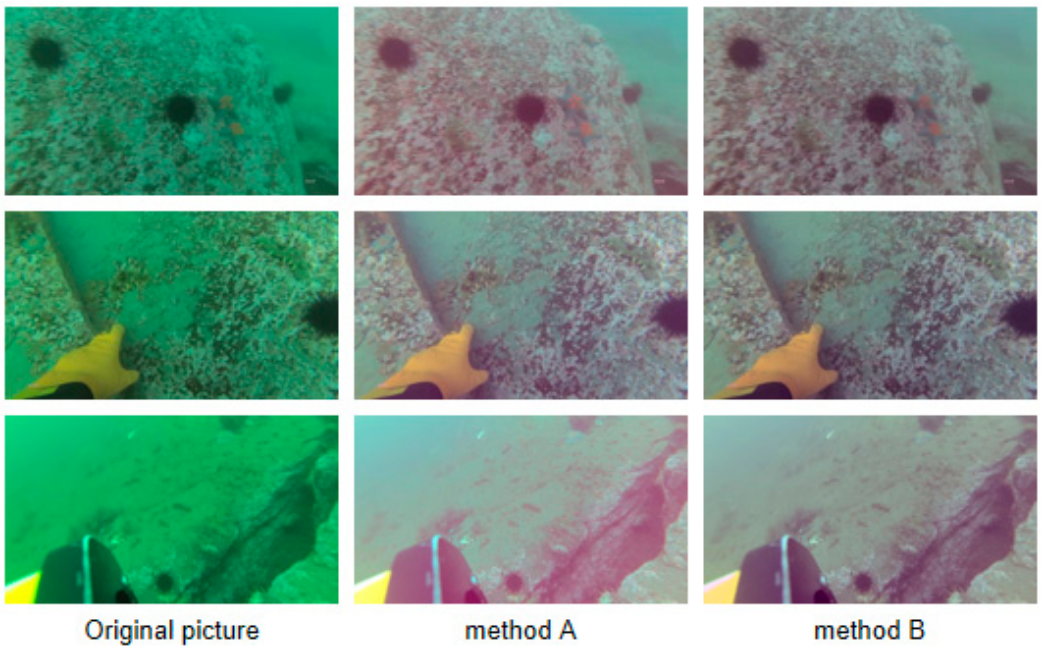


Figure 2. Decolorization of underwater images.

3.2. Deblurring of Underwater Images

After color restoration, it was found that the image was still blurred due to the scattering of light by suspended underwater solids and particles. Therefore, deblurring was carried out next. By referring to the defogging algorithm in the atmospheric environment, the all-in-one dehazing network (AOD) [42] is applied to the underwater environment.

The classical atmospheric scattering model is shown as

$$I(x) = J(x)t(x) + A(x)(1 - t(x)), \tag{12}$$

where $A(x)$ represents the atmospheric light intensity, $t(x)$ represents the transmittance, $I(x)$ represents the fuzzy image, $J(x)$ represents the defogging image.

In order to obtain the defogged image, the above formula can be transformed into

$$J(x) = \frac{I(x)}{t(x)} - \frac{A}{t(x)} + A, \tag{13}$$

where given $I(x)$, only by solving $t(x)$ and A can we obtain $J(x)$. However, this step-by-step solution can easily lead to accumulation errors and sub-optimal solutions. Because of the complexity of the underwater environment, the error will be even greater when using this method to defog underwater images. Therefore, this paper used AOD to deblur the underwater images. The main idea of this method is to unify $t(x)$ and A so that and the optimal solution of $J(x)$ is obtained by directly minimizing the reconstruction error in the pixel domain.

Expression (13) can be rewritten as

$$J(x) = K(x)I(x) - K(x) + b, \tag{14}$$

where

$$K(x) = \frac{\frac{1}{t(x)}(I(\mathrm{x}) - A) + (A - b)}{I(x) - 1}, \tag{15}$$

where b represents a constant deviation value, usually 1.

AOD estimates $K(x)$ through a solution network (adaptive convolution network), and then obtains a deblurred image through Equation (14). The network structure of the AOD solution is shown in Figure 3. The features of different scales are obtained by convolution kernels of different sizes. The middle layer connects the coarse-scale network with the fine-scale network to make the information contained in the feature map more perfect. Conv1~5, respectively, represent five convolution layers, with the middle layer con.cat1 splicing the features of convolution layers 1 and 2, the middle layer con.cat2 splicing the features of convolution layers 2 and 3, and the middle layer con.cat3 splicing the features of convolution layers 1, 2, 3 and 4.

Using the deblurring network, the images in Figure 2 are further processed. The results are shown in Figure 4. Comparing the original image with the two groups of processing results, the images processed by different color cast removal methods have been improved after further deblurring. Comparatively speaking, the color cast image based on the improved white balance algorithm is clearer after deblurring.

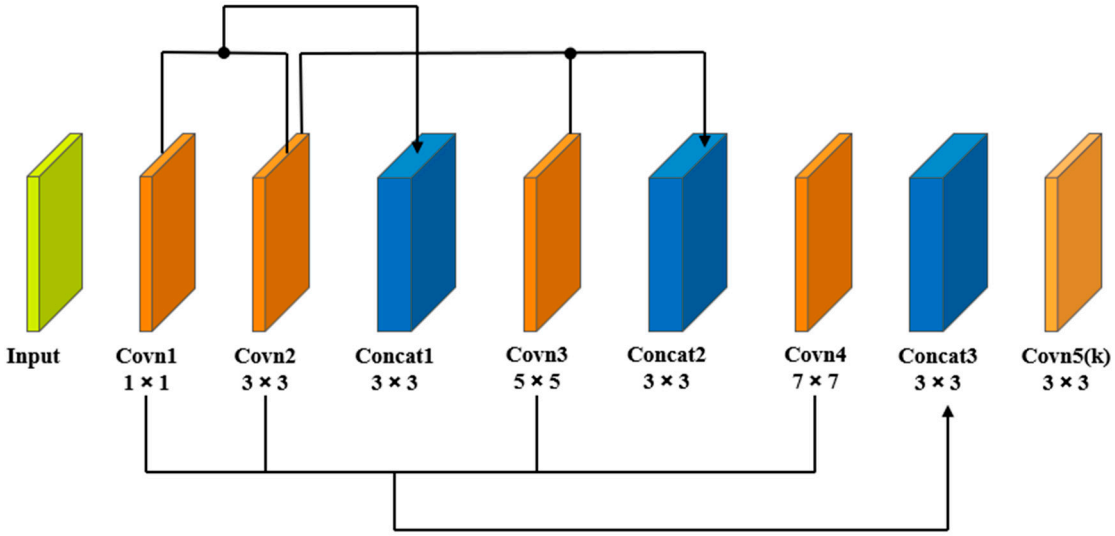


Figure 3. AOD Solving Network Structure.

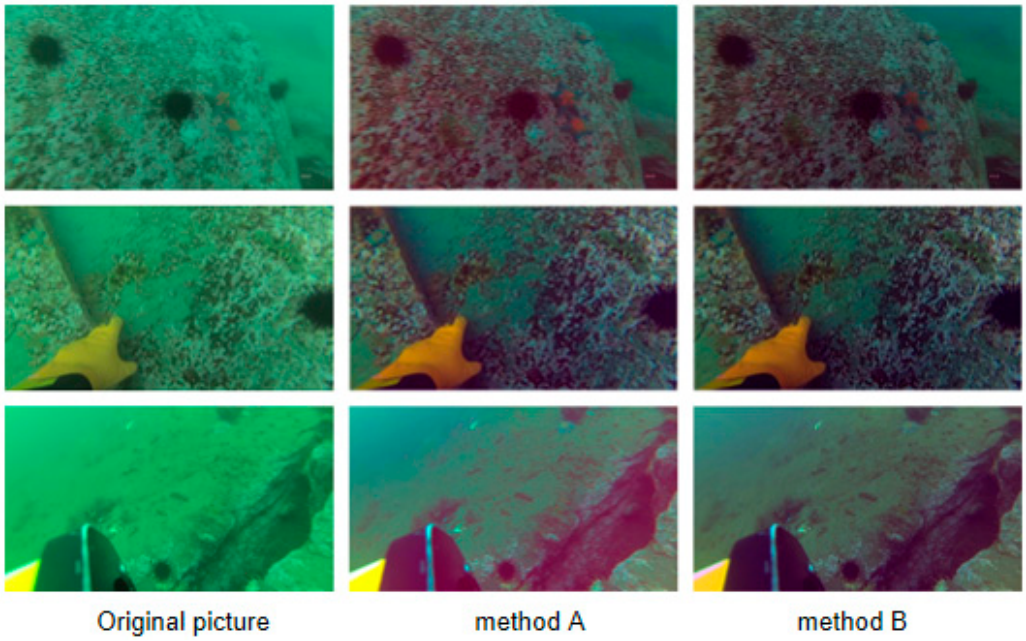


Figure 4. Underwater image deblurring based on AOD and color restoration.

3.3. Contrast Enhancement of Underwater Images

After color shift and deblurring, the underwater images have been obviously enhanced. However, some details in the images are not satisfactory. Considering the characteristics of underwater images, contrast limited adaptive histogram equalization (CLAHE) was used [43]. This method can not only effectively improve the contrast, but also better preserve the detailed information of the image (local texture and edge information, etc.). Using the CLAHE method, the corner region is directly histogram equalized, the boundary region is histogram equalized, and then the pixels in two adjacent regions are linearly interpolated. For the inner region, the pixels in four adjacent regions are bilinear interpolated after histogram equalization [44].

On the basis of deblurring, the CLAHE method was used to further enhance the underwater images. The results are shown in Figure 5. Compared with the original images, it can be seen that both methods have good enhancement effects. Moreover, the processing

based on method A is better on the whole, with clearer details and the more obvious edges of the target.

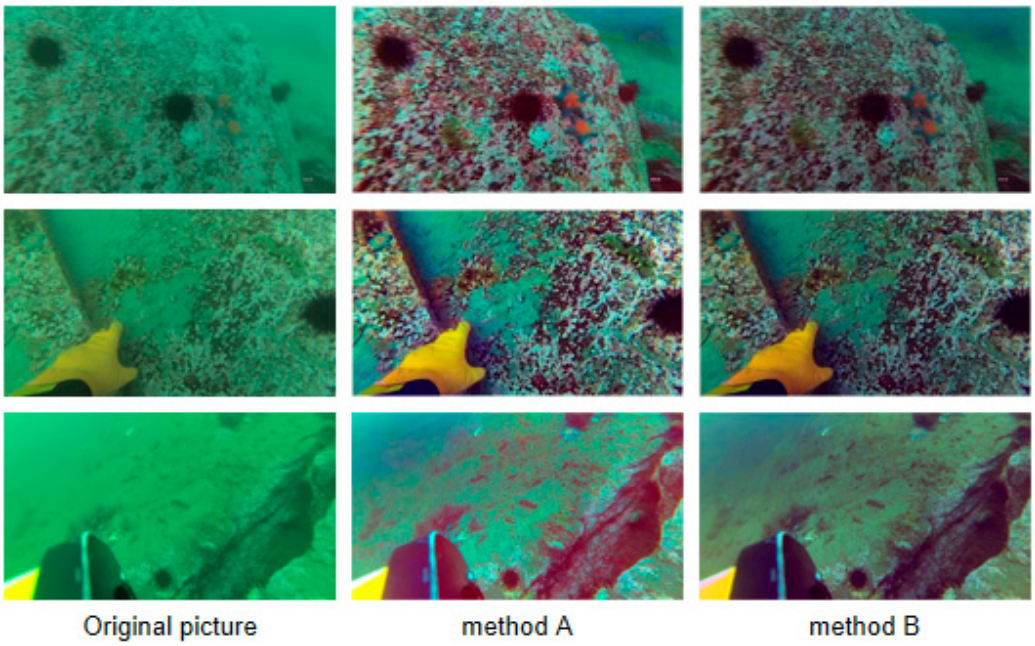


Figure 5. Contrast improvement.

3.4. Underwater Image Enhancement Based on a Combination Algorithm

In order to verify the effectiveness of the proposed combined algorithms, they are compared with the four classical image enhancement algorithms introduced earlier, namely MSRCR, ICM, RGHS, UCM. Table 1 shows the composition of two combination algorithms. The main difference is that the methods of color cast removal are different. For the sake of simplicity, the first combination algorithm is named after WAC while the second combination algorithm is LAC.

Table 1. Two Underwater Image Enhancement Algorithms.

Target	Combination Algorithm WAC	Combination Algorithm LAC
Color cast removal	Improved white balance	Lab color correction
Deblurring	AOD	AOD
Contrast enhancement	CLAHE	CLAHE

Figure 6 shows the effects of six underwater images pretreated by six underwater image enhancement algorithms. Through observation and comparison, it was found that all image enhancement algorithms have certain effects on underwater images, but the effects are different. The MSRCR algorithm has a good enhancement effect when the green color is not significant, but it will be distorted when the green color is significant. The results of ICM and RGHS are similar, which can improve the detailed information of the image, but the effect of color cast removal is average. Compared with the first three algorithms, the UCM algorithm has a better comprehensive enhancement effect on underwater images, and has certain improvement in color cast removal, deblurring and contrast. Compared with the previous algorithms, the two combinatorial optimization algorithms, LAC and WAC, have been further improved, and the texture of the image and the edge information of the target are clearer, and the color cast of the image is better, and the contrast is also better.

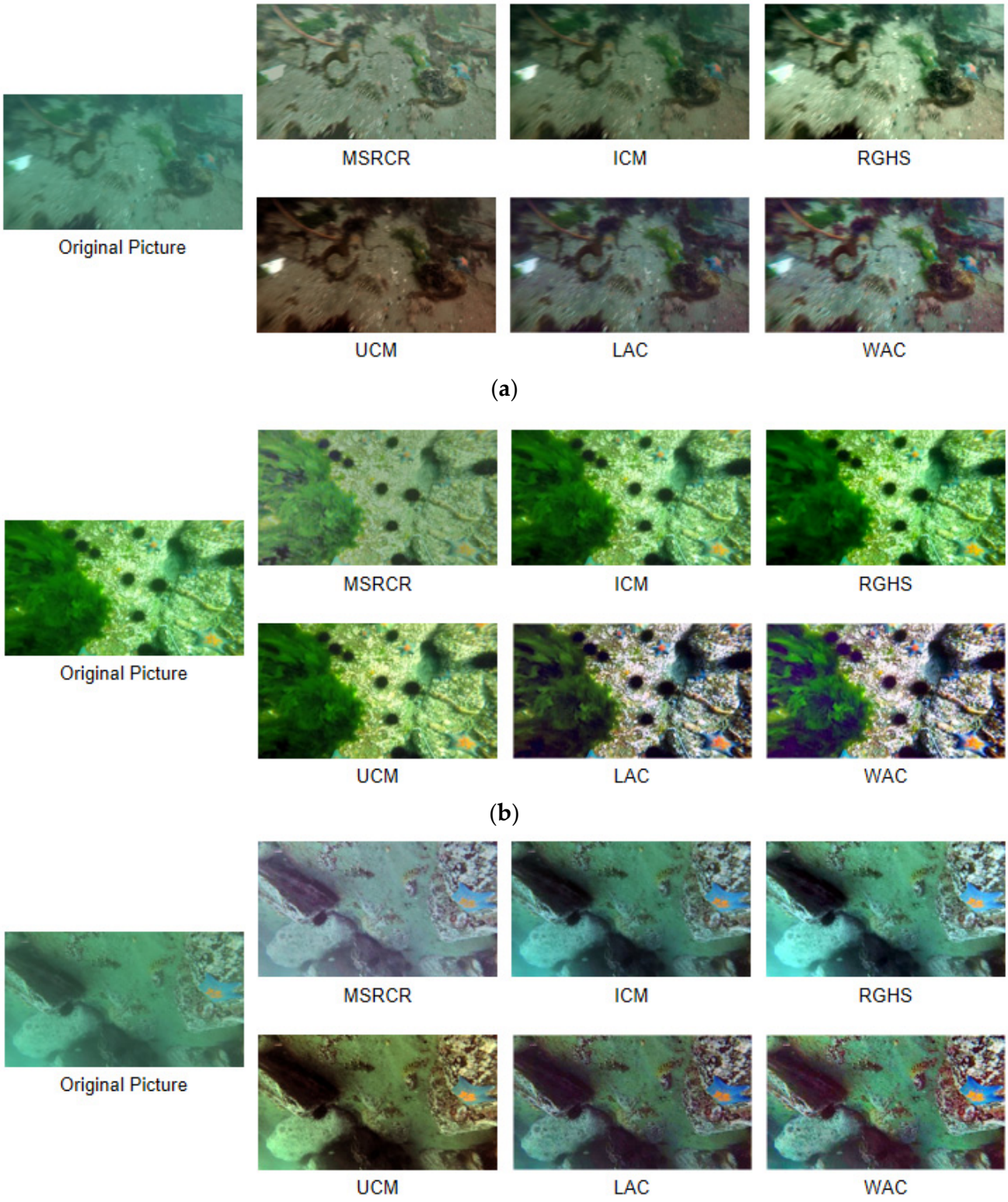


Figure 6. *Cont*.

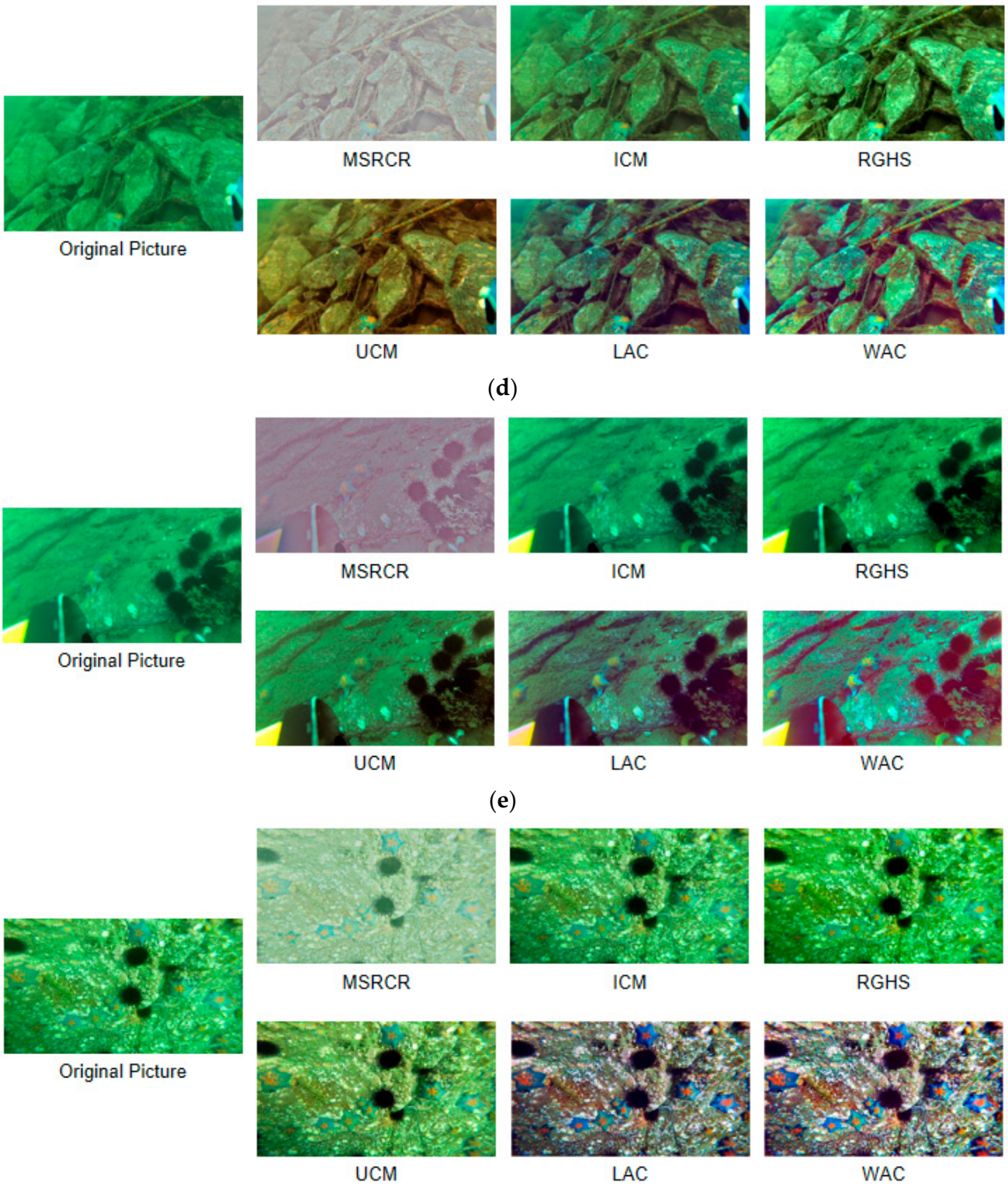


Figure 6. Underwater image enhancement by different algorithms. (**a**) Enhancement effect of No. 1 underwater image; (**b**) Enhancement effect of No. 2 underwater image; (**c**) Enhancement effect of No. 3 underwater image; (**d**) Enhancement effect of No. 4 underwater image; (**e**) Enhancement effect of No. 5 underwater image; (**f**) Enhancement effect of No. 6 underwater image.

Generally speaking, the combination optimization algorithm has a better enhancement effect than several classical algorithms. Compared with the two combined algorithms,

the WAC algorithm has a better enhancement effect than the LAC algorithm. After WAC processing, more detailed information can be seen in the image, the edge information is more obvious, and the color of the image is closer to the real situation.

3.5. Quantitative Evaluation of Underwater Image Enhancement Effect

When evaluating the image enhancement effect, relying solely on visual perception can easily be influenced by individual subjective preferences. To achieve a more objective assessment, it is necessary to introduce some quantitative metrics. Therefore, this paper compares the above algorithms using several evaluation indexes and edge detection.

Evaluation Indicators

In this work, some evaluation indexes were used to evaluate the underwater image after underwater enhancement, including information entropy, root mean square contrast, average gradient and underwater color image quality evaluation (UCIQE).

(a) information entropy

Information entropy can reflect the average amount of information contained in an image, and the greater the entropy, the more information can be obtained from the image [45]. The calculation formula is:

$$Entorpy = -\sum_{i=0}^{n} p(i) log_2 p(i), \tag{16}$$

where, n represents that maximum gray level of the image, i represents the gray level of the pixel, $p(i)$ represents the probability that the pixel value of the image is i.

(b) root mean square contrast

The contrast of an image represents the difference between its gray levels, and the degree of this difference depends on the contrast value of the image. Usually, the greater the value, the higher the quality of the image. At present, the commonly used contrast measurement indicators mainly include root mean square contrast [46], Weber contrast [47] and three kinds of Michelson contrast [48]. The root mean square contrasts calculation formula is:

$$\sigma_{I_{w\times h}} = \sqrt{\frac{1}{w \times h} \sum_{I_{w\times h}} (I(x,y)) - \frac{1}{w \times h} \sum_{I_{w\times h}} I(x,y)^2}, \tag{17}$$

where, w represents the width of the image, h represents the height of the image, $I(x,y)$ represents the gray value of point (x,y) in the image.

(c) average gradient

The average gradient of the image can reflect the clarity and texture changes of the image, and the larger the value, the clearer the image is. The calculation formula is

$$G = \frac{1}{3}\sum_{\lambda} \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}{2}} \lambda \in \{R, G, B\}, \tag{18}$$

where, M represents the width of the image, N represents the height of the image, $\partial f/\partial x$ represents the horizontal gradient of the image, $\partial f/\partial y$ represents the vertical gradient of the image, $f(x,y)$ represents the gray value of point (x,y) of the image.

(d) UCIQE

This index is based on the linear combination of chromaticity, contrast and saturation of the image under the Lab color model [49]. The larger the value, the better the image quality. The calculation formula is:

$$G = UCIQE = C_1 \times \sigma_c + C_2 \times con_1 + C_3 \times \mu_s, \quad (19)$$

where, C_1, C_2 and C_3 represent the weight coefficient, σ_c represents the labeling deviation of chromaticity, con_1 represents the contrast value of brightness, μ_s represents the average value of saturation.

The underwater image enhancement results in Figure 6 are evaluated by using the above four evaluation indexes, and the results are shown in Tables 2–5.

Table 2. Evaluation Results of Information Entropy.

	Original Picture	MSRCR	ICM	RGHS	UCM	LAC	WAC
No. 1	6.2183	6.5914	6.7442	7.6782	6.8203	6.8411	7.2714
No. 2	7.5843	6.9128	7.6918	7.7484	7.7663	7.7946	7.8584
No. 3	6.6885	6.9248	7.3297	7.6937	7.5886	7.1999	7.4996
No. 4	6.7032	5.8582	6.9670	7.6812	7.2682	7.2923	7.5383
No. 5	6.8069	5.6909	6.9863	7.2281	7.2352	7.3048	7.3921
No. 6	7.3638	6.3914	7.4306	7.5637	7.5537	7.7738	7.9223

Table 3. Evaluation Results of Root Mean Square Contrast.

	Original Picture	MSRCR	ICM	RGHS	UCM	LAC	WAC
No. 1	1.5174	3.2502	2.1309	3.9588	2.4760	3.9923	5.3101
No. 2	9.0783	6.4168	9.5961	10.7941	10.1105	16.5052	17.1660
No. 3	4.5149	7.4009	6.9891	9.0983	8.8501	10.2568	13.3756
No. 4	2.7951	3.1681	3.2853	5.4359	4.6006	6.1526	8.3490
No. 5	3.1137	3.5767	3.3553	4.3159	4.2166	6.3078	9.3100
No. 6	11.3985	7.9082	11.8319	13.3216	12.8463	20.8758	24.5773

Table 4. Evaluation Results of Average Gradient.

	Original Picture	MSRCR	ICM	RGHS	UCM	LAC	WAC
No. 1	0.8266	2.0057	1.2266	2.3023	1.4063	2.1515	3.0642
No. 2	5.1933	4.2321	5.5190	6.0015	5.8072	9.5737	10.2269
No. 3	2.5784	4.4454	4.0649	5.2726	5.0802	6.1491	8.0715
No. 4	1.4006	1.6943	1.6975	2.7983	2.3888	3.2234	4.5861
No. 5	1.5935	1.9307	1.7326	2.2795	2.2322	3.3477	4.9944
No. 6	6.3686	4.4176	6.6017	7.4911	7.1711	11.5833	13.7102

Table 5. Evaluation Results of UCIQE.

	Original Picture	MSRCR	ICM	RGHS	UCM	LAC	WAC
No. 1	0.2752	0.2213	0.3296	0.3897	0.3693	0.3625	0.3990
No. 2	0.4188	0.2796	0.4361	0.4769	0.4471	0.4613	0.4817
No. 3	0.2736	0.2615	0.3775	0.4322	0.4426	0.3962	0.4400
No. 4	0.3366	0.1994	0.3529	0.4369	0.4419	0.4343	0.4640
No. 5	0.4009	0.1605	0.4240	0.5008	0.4324	0.4511	0.4568
No. 6	0.4015	0.2160	0.4138	0.4662	0.4397	0.4705	0.4907

As can be seen, the proposed combination algorithm, WAC, is generally superior to other algorithms. In terms of the information entropy index, WAC based evaluation results in images 1, 3 and 4 are slightly worse than the RGHS algorithm. Nevertheless, WAC is better than the other four algorithms. Under the two indexes of root mean square contrast and average gradient, the WAC and LAC proposed in this work are obviously better than the other algorithms, and WAC has the best effect. In terms of UCIQE, the WAC algorithm achieved better results except that the evaluation of image No. 5 is slightly worse than the RGHS algorithm. In summary, we can find that the combination algorithm proposed has a good enhancement effect on underwater images. Compared with other algorithms, the combined algorithm is more effective in color correction, contrast improvement and information increase, among which the WAC algorithm is generally better than the other algorithms.

4. Underwater Image Instance Segmentation Based on an Improved Mask R-CNN Network

On the basis of data set augmentation and image enhancement, this paper uses a Mask R-CNN network to perform underwater image segmentation. In order to improve the accuracy, some improvements to the structure of the Mask R-CNN network are conducted, mainly based on the characteristics of the data sets to make some improvements to the FPN network.

4.1. Mask R-CNN Network

The Mask R-CNN is a very flexible framework, which can accomplish different tasks by adding different branches. The general framework of mask-RCNN is the framework of faster-RCNN. By adding a fully connected branch of Mask, three tasks of classification, regression and segmentation are performed. Mask R-CNN is a two-stage detector: in the first stage, the feature map is scanned through the region proposal network (RPN) and proposals are generated, that is, the region that may contain a target; in the second stage, RoIAlign is used to extract features from each candidate frame, classify and regress the bounding box, and a full convolution network branch is added to predict and output the corresponding binary mask for each region of interest. Mask-RCNN is a two-stage model [50]. In the first stage, known as the region proposal network (RPN), the initial feature maps are scanned to generate region proposals or regions of interest (RoI). This process is similar to the faster-RCNN model. In the second stage, RoI-pooling is applied to each RoI to downsample the feature map using a nearest neighbor approach. This step helps select important features from the feature map. However, RoI-pooling can lead to misalignment between the RoI and the extracted features. To address this issue, RoI-align is applied to each RoI to create more accurate RoIs. In RoI-align, the value of each sample point is calculated using bilinear interpolation from nearby grid points on the feature map. Additionally, Mask-RCNN not only predicts the class and bounding boxes for each object but also generates a binary mask for each RoI using a fully convolutional network (FCN). Figure 7a provides an overview of the Mask-RCNN network framework.

The structure diagram of the FPN network is shown in Figure 7b, in which T2–T5 are the characteristic diagrams of different levels output by the backbone network, and the size relationship between T2 and T5 is twice that of the upper layer, so the lower layer is spliced with the upper layer's characteristic diagram through upsampling, and then different characteristic diagrams P2–P5 are output through convolution, and P6 is obtained through maximum pooling of P5. These feature maps of different levels contain information of different scales in the original map, which is exactly the function of the FPN network. P2–P5 are used for RPN networks and classified network parts, while P6 is only used for RPN networks. In this work, the improvement of the FPN network is the P6 part.

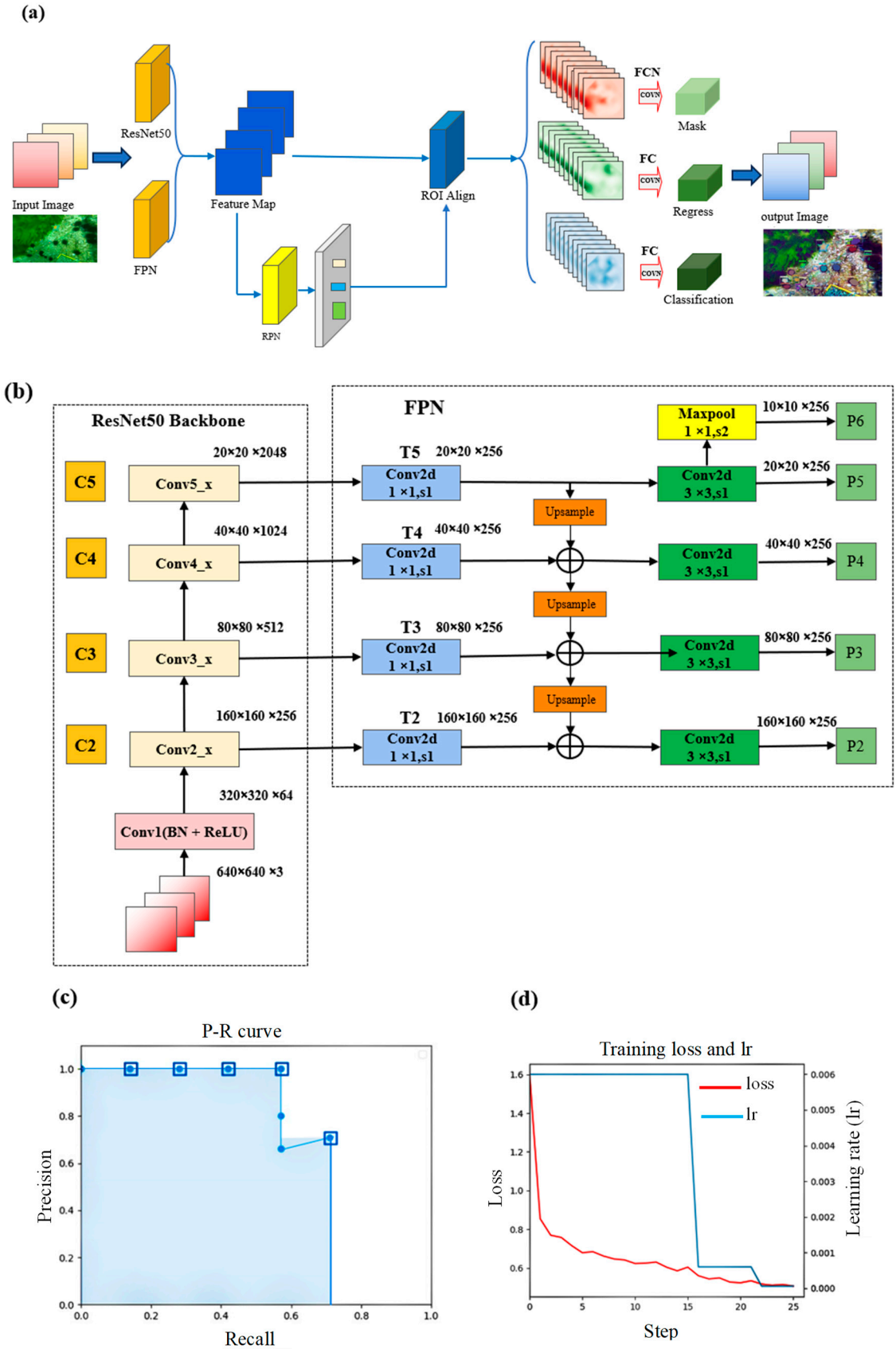


Figure 7. Mask R-CNN network structure. (**a**) Mask R-CNN network model. (**b**) ResNet50 Backbone and FPN network structure improvement. (**c**) Evaluation index of underwater image instance segmentation, PR curve. (**d**) Network training method.

Based on the characteristics of this data set, the target size in the image is relatively small and the depth of the backbone network is relatively large. Because the receptive field of the output feature map is relatively large, which is not conducive to the training of the network and the final segmentation accuracy, this paper directly removes the P6 part of the FPN network, which is only used for the RPN network. This can also reduce the parameters of the network and speed up the training of the network.

Usually, the index for evaluating the accuracy of image segmentation and target recognition is the mean average precision (mAP), which represents the average value of each category accuracy AP. AP is obtained according to PR curve, as shown in Figure 7c. The ordinate represents the precision rate p of a class, and the abscissa represents the recall rate r of a class. The figure shows five groups of values of a class, and the area enclosed by the PR curve and the coordinate axis (blue part) is 0.67, so the AP value of this class is 0.67.

Based on the sample size of this data set, the network adopts the methods of migration training and freezing to train. Before network training, the pre-training weights of Resnet50 and Mask R-CNN on COCO data set are loaded for migration learning. Doing so can speed up the convergence of the network and achieve better training results. Other training parameters are as follows: the epoch is set to 26, and the learning rate is set to 0.006 (the learning rate will decrease at the 16th epoch with the increase of the epoch), and the optimizer adopts random gradient descent (SGD) [51]. As shown in Figure 7d, the changes of training loss and learning rate with epoch (taking the data set after data augmentation as an example) show that the network converges in a relatively fast time.

4.2. Experimental Results and Analysis

4.2.1. Validation of Data Augmentation Method

Experiments were carried out on the initial data set and the data set after data augmentation (image rotation, image flipping and images generated by ConSinGan) to verify the effectiveness of data augmentation in accuracy. The results are shown in Table 6, where S stands for recognition, B stands for segmentation, mAP stands for average accuracy, mAR stands for average recall, 0.5 stands for IOU threshold value of 0.5, and all stands for the IOU threshold value of 0.5 to 0.95.

Table 6. Experimental Results with or without Data Enhancement.

Data Augmentation	mAP[all]/s	mAP[all]/b	mAP[0.5]/b	mAR[all]/s	mAR[all]/b
with	0.571	0.126	0.275	0.651	0.247
without	0.517	0.123	0.239	0.612	0.198

Comparing the experimental results of the two groups, it can be seen that after data augmentation, all the indexes of mAP and mAR are improved steadily, so the effectiveness of the data augmentation method we proposed is verified.

4.2.2. Validation of Image Enhancement Algorithm

Four image enhancement algorithms (MSRCR, ICM, RGHS, UCM) and the two combinatorial optimization algorithms (WAC and LAC) proposed in this work were used to preprocess the data set, and the accuracy of the preprocessing method was verified. The results are shown in Table 7.

Comparing several groups of experimental results, it can be seen that the recognition effect is similar to that without preprocessing except for MSRCR algorithm. On the segmentation effect, all of them have been improved after preprocessing, which shows that the preprocessing method is effective for segmentation. The order of the improvement effect is MSRCR, ICM and UCM are similar, which is further improved than MSRCR. The combination optimization method LAC is better than RGHS, and WAC is the best. Therefore, the effectiveness of the combined optimization algorithm proposed in this paper in underwater image enhancement is verified.

Table 7. Experimental Results of Image Enhancement Algorithm.

Pretreatment Method	mAP[all]/s	mAP[all]/b	mAP[0.5]/b	mAR[all]/s	mAR[all]/b
without	0.571	0.126	0.275	0.651	0.247
MSRCR	0.461	0.134	0.292	0.585	0.231
ICM	0.563	0.205	0.376	0.634	0.308
RGHS	0.559	0.217	0.424	0.647	0.328
UCM	0.575	0.181	0.364	0.656	0.280
LAC	0.577	0.225	0.448	0.653	0.336
WAC	0.573	0.234	0.456	0.650	0.349

4.2.3. Validation of the Improvement of Mask R-CNN Network

The improved FPN network is compared and tested. The results are shown in Table 8.

Table 8. Experimental Results of Network Improvement.

Improve	mAP[all]/s	mAP[all]/b	mAP[0.5]/b	mAR[all]/s	mAR[all]/b
without	0.571	0.126	0.275	0.651	0.247
FPN reform	0.571	0.157	0.300	0.652	0.254

Comparing the experimental results of three groups, it can be seen that the recognition results of the improved FPN network are similar, but the segmentation effect has been improved to some extent, and the speed of image processing by the network has also been slightly improved due to the reduction of network parameters, thus verifying the effectiveness of the modification of the FPN network. In addition, after the backbone network is changed to MobilenetV2, the recognition and segmentation effect of the network is slightly reduced, but the processing speed is greatly improved.

4.2.4. Comprehensive Verification

Through the previous experiments, the effectiveness of the image enhancement algorithm (image preprocessing algorithm) and Mask R-CNN network improvement we proposed was verified, and the results are shown in Table 9, where FL and FW, respectively represent the combination of FPN network improvement with LAC and WAC. The model participating in the comparison is a single-stage instance segmentation network Yolact [52].

Table 9. Comprehensive Verification Experimental Results.

Model	mAP[all]/s	mAP[all]/b	mAP[0.5]/b	mAR[all]/s	mAR[all]/b
Initial	0.571	0.126	0.275	0.651	0.247
FL	0.586	0.231	0.453	0.672	0.339
FW	0.591	0.245	0.483	0.664	0.352
Yolact	0.425	0.116	0.267	0.513	0.226

Compared with several groups of experimental results, it can be seen that the combination of the image preprocessing method and network improvement further improves the effect of recognition and segmentation, and the accuracy is steadily improved on the basis of only preprocessing and network improvement, thus verifying the effectiveness of combining them together. In addition, another example is used to segment the network, Yolact. Compared with this method, the proposed method has better recognition and segmentation effects, which verifies the effectiveness of the proposed method.

Figures 8–13 illustrate the optimal outcomes of underwater image segmentation by using the FW method combined with six underwater images in which three kinds of underwater creatures, i.e., echinus, starfish and holothurian, are to be segmented. The number of targets to be segmented increases from the No. 1 image to the No. 6 image. The percentages in the graph are the probabilities of the target being recognized as the

corresponding category. As can be seen, the probabilities are generally above 80%, which indicates that the proposed FW method can well achieve the segmentation of underwater image instances. In detail, in Figures 8, 10 and 11, the proposed FW method did not miss any targets and segmented the targets well. Figures 12 and 13 show that the segmentation of underwater images containing many targets has also achieved good results. However, it should be noted that the segmentation effect is insufficient for incomplete targets at the edge of the image. As can be seen, one starfish at the bottom is missed for segmentation in Figures 9 and 12, shown in dashed-boxes. In addition, the underwater environment is complex and diverse, which has an impact on the segmentation and recognition of underwater targets. For example, in Figure 12, the recognition probability of a starfish is only 53%, well below the average level. In Figure 13, two starfishes and several possible echini are missed for segmentation. Nevertheless, on the whole, the proposed method achieves the instance segmentation of underwater images well.



Figure 8. Segmentation result of No. 1 image by using FW method, all segmented.

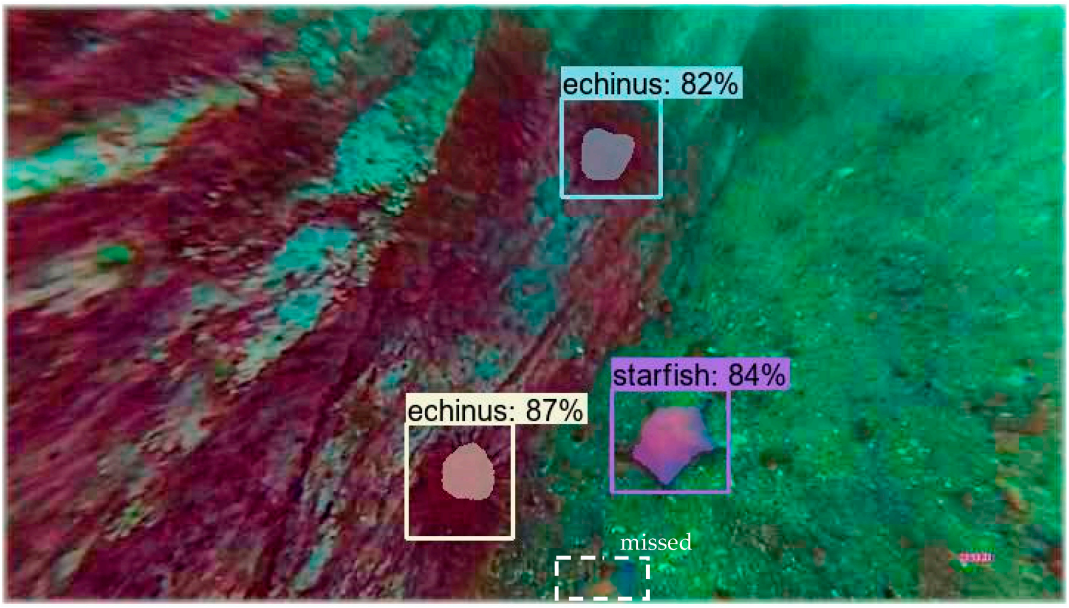


Figure 9. Segmentation result of No. 1 image by using FW method, edged target missed.

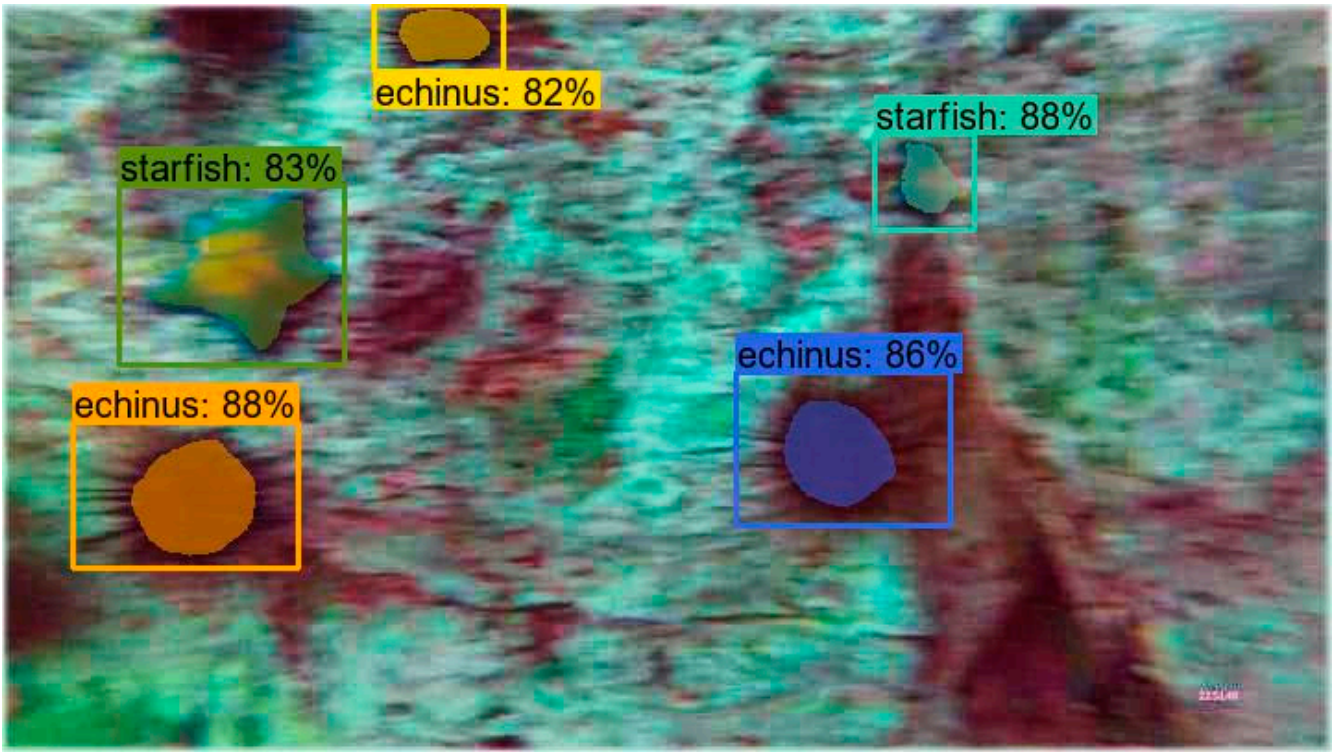


Figure 10. Segmentation result of No. 3 image by using FW method, all segmented.

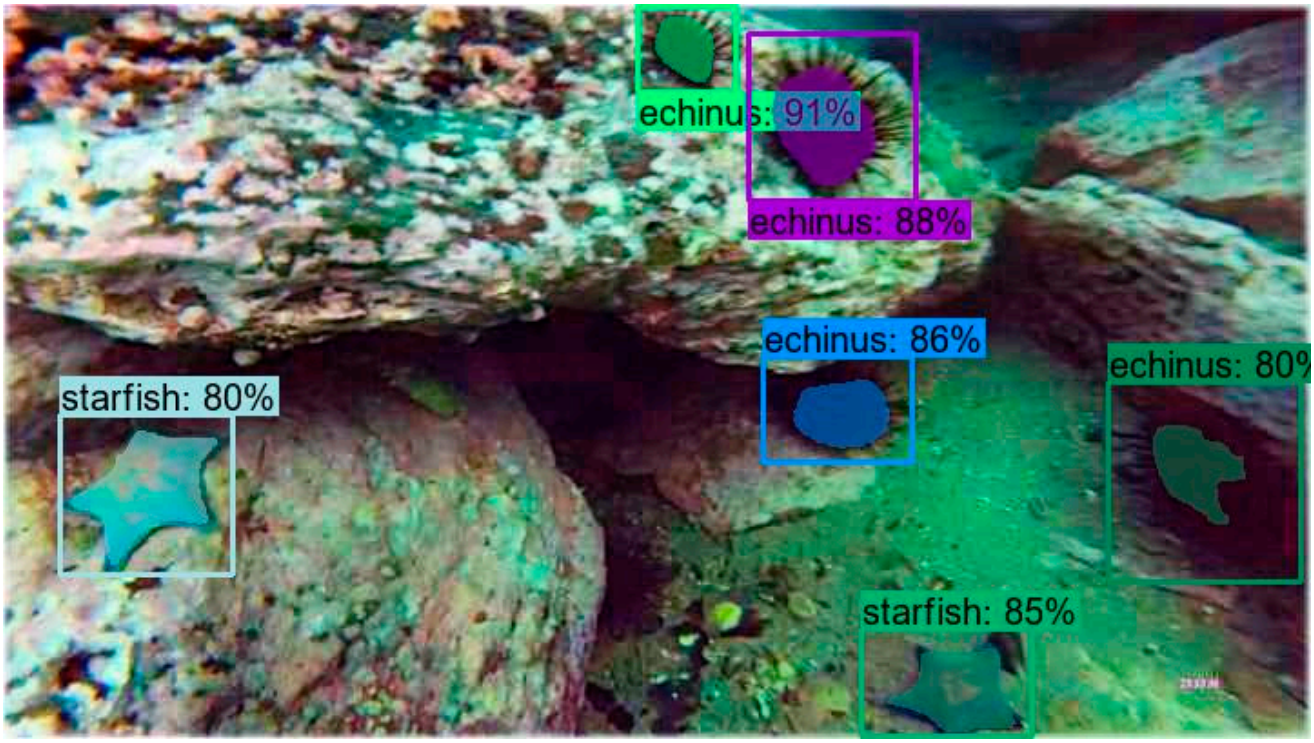


Figure 11. Segmentation result of No. 4 image by using FW method, all segmented.

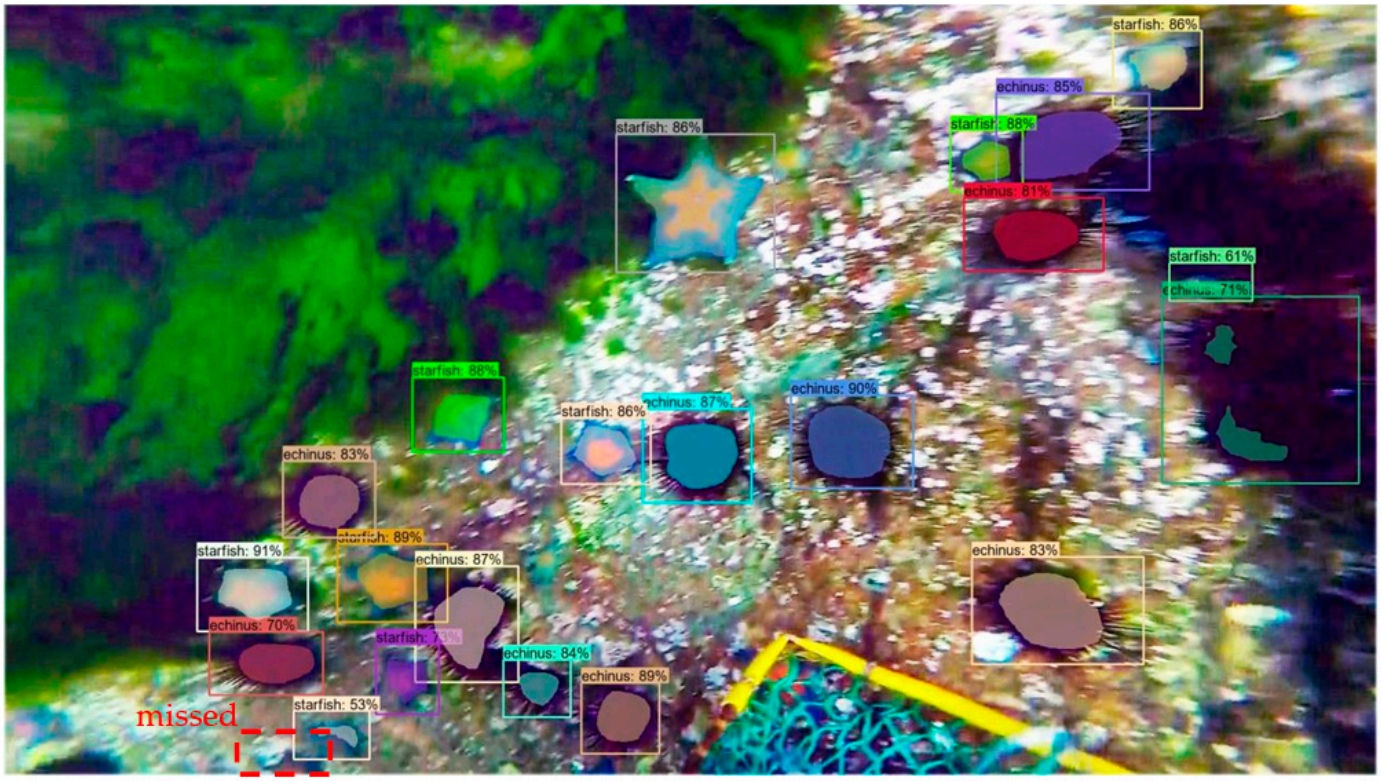


Figure 12. Segmentation result of No. 5 image by using FW method, edged target missed.

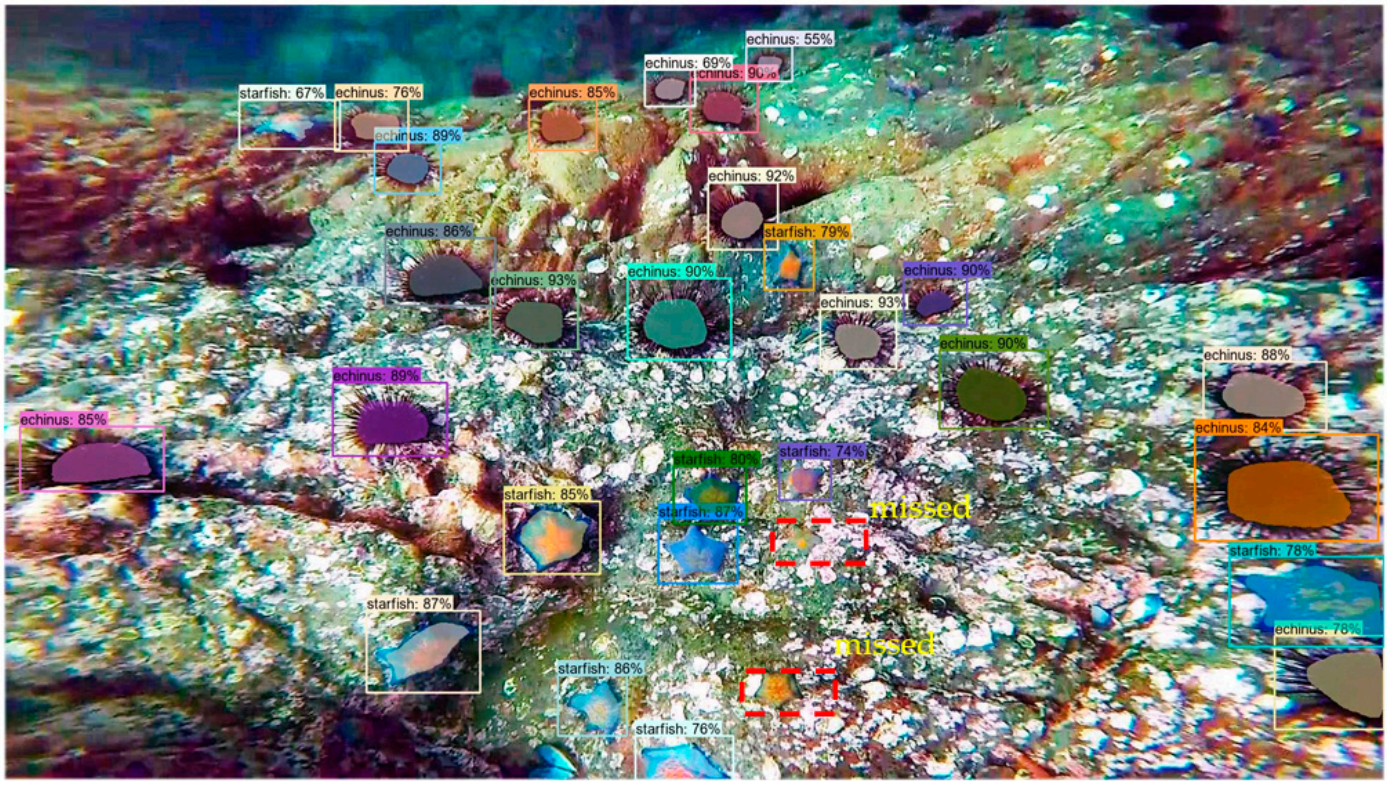


Figure 13. Segmentation result of No. 6 image by using FW method, background targets missed.

5. Conclusions

In this research, we demonstrate that deep neural network and image enhancement combined optimization algorithm can significantly improve segmentation of average accuracy. In view of the lack of high-quality underwater data sets, the data sample size is expanded by data augmentation, including image rotation and flipping, and images generated by a GAN network. We can achieve an mAP (segmentation of average accuracy) of 0.126. Aiming at the problems of color deviation, blur and poor contrast of underwater images caused by the complex underwater environment and imaging characteristics, an image enhancement algorithm can be used for preprocessing to improve the quality of the data sets. The algorithms include multi-scale Retinex (MSRCR) with color restoration, integrated color model (ICM), relative global histogram stretching (RGHS), and unsupervised color correction (UCM), as well as the combined optimization algorithms (WAC and LAC) that we proposed, which are carried out step by step in three aspects: color cast removal, deblurring and contrast enhancement. Through subjective visual judgment and an objective evaluation index, it was verified that the preprocessing method is effective and that the combined optimization algorithm that we proposed has a better effect than other algorithms. Specifically, compared with the conventional algorithm and state-of-the-art image enhancement algorithm, this combined optimization algorithm can correspondingly improve the mAP (segmentation of average accuracy) speed up to 0.1~0.9 times. It can largely increase the segmentation of average accuracy. In addition, based on the characteristics of constructing underwater image data sets, the feature pyramid network (FPN) was improved to some extent, and the experimental results show that the improvement of the network improves the segmentation and recognition accuracy to some extent. Furthermore, the preprocessing method was combined with the improved network to carry out experiments and compared with other neural networks to verify the effectiveness of the proposed method, and the effect and purpose of improving underwater image instance segmentation and target recognition were achieved. The results demonstrate that our method (FW) can significantly increase the segmentation of average accuracy by 1.1 times compared with the conventional target recognition models.

Due to the difficulty and time-consuming production of high-quality underwater image datasets, the datasets constructed in this work have fewer categories. Therefore, it is necessary to further verify the efficiency of this method in target recognition and segmentation of other marine biological categories. Moreover, it should be noted that the method proposed in this work is a combination of optimization algorithms, which implies its long processing time. In future work, we intend to improve the real-time performance of the algorithm so that it may be better applied in underwater image target recognition and segmentation. In addition, we will use other optimized neural network models combined with larger underwater image datasets to further improve the accuracy of the model.

Author Contributions: Conceptualization, J.C. and S.Z.; methodology, J.C. and S.Z.; software, J.C. and S.Z.; validation, W.L.; formal analysis, J.C. and S.Z.; investigation, J.C. and S.Z.; resources, W.L.; data curation, J.C. and S.Z.; writing—original draft preparation, J.C. and S.Z.; writing—review and editing, W.L.; visualization, J.C. and S.Z.; supervision, W.L.; project administration, W.L.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fuzhou Institute of Oceanography, Grant 2022F13.

Data Availability Statement: The data can be shared up on request.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive suggestions which comprehensively improve the quality of the paper.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Wang, J. Research on Real-Time Underwater Target Recognition Algorithm Based on Image Restoration and YOLO. Master's Thesis, Harbin University of Science and Technology, Harbin, China, 2021.
2. Ahn, J.; Yasukawa, S.; Sonoda, T.; Ura, T.; Ishii, K. Enhancement of deep-sea floor images obtained by an underwater vehicle and its evaluation by crab recognition. *J. Mar. Sci. Technol.* **2017**, *22*, 758–770. [CrossRef]
3. Singh, H.; Adams, J.; Mindell, D.; Foley, B. Imaging underwater for archaeology. *J. Field Archaeol.* **2000**, *27*, 319–328.
4. Watanabe, J.-I.; Shao, Y.; Miura, N. Underwater and airborne monitoring of marine ecosystems and debris. *J. Appl. Remote Sens.* **2019**, *13*, 044509. [CrossRef]
5. Lin, N.; Zhang, D.; Zhang, K.; Wang, S.; Fu, C.; Zhang, J.; Zhang, C. Small sample convolution neural network learning and prediction of seismic oil and gas reservoirs. *J. Geophys.* **2018**, *61*, 4110–4125.
6. Dai, H.; Lei, F.; Shang, S.; Lin, R.; He, Z. A Meaningful Wave Height Prediction Method Based on Deep Learning. CN109460874A, 12 March 2019.
7. Gou, J.; Jiang, Y.; Li, Z.; Dong, X.; Wu, B.; Liu, Z. Information Extraction of Aquaculture Water Body in Chengdu Plain Based on Deeplapv3+ Model. *J. Agric. Mach. Chem. China* **2021**, *42*, 105–112.
8. Yang, M.; Lei, B.; Zhao, Q.; Lan, R. Two-dimensional fuzzy divergence multi-threshold image segmentation based on improved particle swarm optimization. *Comput. Appl. Softw.* **2020**, *37*, 133–138.
9. Liu, Y.; Wang, Y.; Yu, H.; Qin, M.; Sun, J. Detection of straw coverage based on multi-threshold image segmentation algorithm. *J. Agric. Mach.* **2018**, *49*, 27–35.
10. Ma, L.; Zhang, Y.; Deng, J. Watershed algorithm based on morphological opening and closing filtering binary mark and texture feature merging. *J. Image Graph.* **2003**, *8*, 80–86.
11. Mariena, A.A.; Sathiaseelan JG, R.; A braham, J.T. Hybrid approach for image segmentation using region splitting and clustering techniques. In Proceedings of the 2018 IEEE International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kerala, India, 21–22 December 2018; pp. 1–4.
12. Cai, Y.; Jia, Z. A new image segmentation algorithm based on fuzzy C-means clustering and spatial information. *Laser J.* **2009**, *49–50*, 52.
13. Wang, S.; Xu, Y.; Wan, L.; Tang, X. A fast FCM clustering underwater image segmentation algorithm based on information entropy constraint. *Comput. Sci.* **2010**, *37*, 243–246.
14. Sun, Y.; Chen, Z.; Wang, H.; Zhang, Z.; Shen, J. Level Set Underwater Image Segmentation Based on Region and Edge Features. *J. Image Graph.* **2020**, *25*, 824–835.
15. Li, T.; Tang, X.; Pang, Y. Underwater image segmentation based on improved particle swarm optimization algorithm and fuzzy entropy. *Ocean Eng.* **2010**, *28*, 128–133.
16. Yan, M.; Huang, B.; Zhu, D. Underwater image segmentation based on gray fluctuation. *J. Harbin Eng. Univ.* **2020**, *41*, 1268–1273.
17. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 640–651.
18. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
19. Yu, F.; Koltun, V.; Funkhouser, T. Dilated residual networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattem Reco Gnition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 472–480.
20. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
21. Dai, J.; He, K.; Li, Y.; Ren, S.; Sun, J. Instance-sensitive fully convolutional networks. In Proceedings of the 2016 Springer European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 534–549.
22. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Coference on Computer Version (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
24. Cao, J.; Han, F.; Wang, M.; Zhuang, Y.; Zhu, Y.; Zhang, Y. Underwater image multi-objective semantic segmentation algorithm based on YOLOv5 and FCN-DenseNet. *Comput. Syst. Appl.* **2022**, *31*, 309–315.
25. Chen, H.; Liu, Y.; Liu, C. Underwater semantic segmentation with less tags with information interaction. *J. Dalian Marit. Univ.* **2022**, *48*, 95–102.
26. Yue, Y.; Geng, L.; Zhao, H.; Wang, H. Research on underwater fish image segmentation algorithm based on ARD-PSPNet network. *Optoelectronics* **2022**, *33*, 1173–1182.
27. Hu, X.; Yan, T. Detection algorithm of sea cucumber and starfish based on improved Mask R-CNN. *J. China Metrol. Univ.* **2023**, *34*, 34–43.
28. Hinz, T.; Fisher, M.; Wang, O.; Wermter, S. Improved Techniques for Training Single-Image GANs. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision, Online, 3–8 January 2021; pp. 1299–1308.

29. Sisman, B.; Vijayan, K.; Dong, M.; Li, H. SINGAN: Singing voicc conversion with gencrative adversarial networks. In Proceedings of the 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 18–21 November 2019; pp. 112–118.
30. Tabassum, S.; Ullah, S.; Al-Nur, N.H.; Shatabda, S. Poribohon-BD: Bangladeshi local vehicle image dataset with annotation for classification. *Data Brief* **2020**, *33*, 106465. [CrossRef] [PubMed]
31. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
32. Land, E.H. The Retinex Theory of Color Vision. *Sci. Am.* **1978**, *237*, 108–128. [CrossRef] [PubMed]
33. Bindhu, A.; Uma, M.O. Color Corrected Single Scale Retinex Based Haze Removal and Color Correction for Underwater Images. *Color Res. Appl.* **2020**, *45*, 1084–1903.
34. Barnard, B.K. Investigations into Multi-Scale Retinex. *Color Imaging Multimed.* **1999**, *98*, 9–17.
35. Rahman, Z.U.; Woodell, G.A. Multi-scale Retinex for color image enhancement. In Proceedings of the International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996; pp. 1003–1006.
36. Iqbal, K.; Salam, R.A.; Osman, A.; Talib, A.Z. Underwater Image Enhancement Using an Integrated Colour Model. *IAENG Int. J. Comput. Sci.* **2007**, *34*, 239–244.
37. Huang, D.; Wang, Y.; Song, W.; Sequeira, J.; Mavromatis, S. Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In Proceedings of the 24th International Conference on MultiMedia Modeling (MMM), Bangkok, Thailand, 5–7 February 2018; pp. 453–465.
38. Iqbal, K.; Odetayo, M.; James, A.; Salam, R.A.; Talib, A.Z.H. Enhancing the low quality images using unsupervised colour correction method. In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
39. Luo, W.; Duan, S.; Zheng, J. Underwater image restoration and enhancement based on a fusion algorithm with color balance, contrast optimization and histogram stretching. *IEEE Access* **2021**, *9*, 31792–31804. [CrossRef]
40. Zheng, M.; Luo, W. Underwater Image Enhancement Using Improved CNN Based Defogging. *Electronics* **2022**, *11*, 150. [CrossRef]
41. Kaur, A.; Kranthi, B.V. Comparison between YCbCr color space and CIELab color space for skin color segmentation. *Int. J. Appl. Inf. Syst.* **2012**, *3*, 30–33.
42. Zhu, S.; Luo, W.; Duan, S. Enhancement of Underwater Images by CNN-Based Color Balance and Dehazing. *Electronics* **2022**, *11*, 2537. [CrossRef]
43. Reza, A.M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2004**, *38*, 35–44. [CrossRef]
44. Zimmerman, J.B.; Pizer, S.M.; Staab, E.V.; Perry, J.R.; McCartney, W.; Brenton, B.C. An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. *IEEE Trans. Med. Imaging* **1988**, *7*, 304–312. [CrossRef] [PubMed]
45. Tsai, D.Y.; Lee, Y.; Matsuyama, E. Information entropy measure for evaluation of image quality. *J. Digit. Imaging* **2008**, *21*, 338–347.
46. Yang, M.; Sowmya, A. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [CrossRef]
47. Peli, E. Contrast in complex images. *JOSA A* **1990**, *7*, 2032–2040. [CrossRef]
48. Michelson, A.A. *Studies in Optics*; Courier Corporation: Chelmsford, MA, USA, 1995; pp. 10–46.
49. Schreiber, W.F. *Fundamentals of Electronic Imaging Systems: Some Aspects of Image Processing*; Springer Science & Business Media: Berlin, Germany, 2012; pp. 84–103.
50. Lin, Y.; Min, Y.; He, Z.; Cao, H.; Wang, J. Optimization and implementation of Canny image edge detection based on MATLAB. *Mod. Inf. Technol.* **2022**, *6*, 81–84.
51. Shi, J.; Wang, D.; Shang, F.; Zhang, H. Research progress of stochastic gradient descent algorithm. *Acta Autom. Sin.* **2021**, *47*, 2103–2109.
52. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 27 October–2 November 2019; pp. 9156–9165.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.